2024-2025 PREMIER LEAGUE DATA ANALYSIS PROJECT:

Objective:

*The objective of this project is to analyze* ***team and player performance in the 2024/25 Premier League season*** *using advanced football metrics such as* ***expected goals (xG), expected assists (xA), and possession-based statistics****. By comparing actual outcomes with expected performance indicators, the project aims to:*

* *Identify teams that are* ***overperforming or underperforming*** *relative to expectations.*
* *Evaluate* ***player contributions*** *beyond traditional goals and assists.*
* *Highlight* ***key performance indicators (KPIs)*** *that most strongly influence match outcomes.*
* *Provide* ***data-driven insights*** *that can support fan discussions, tactical analysis, and potential use cases in Fantasy Premier League (FPL).*

*To evaluate and compare the efficiency and effectiveness of Premier League teams and players in the 2024/25 season using data analytics and advanced football performance metrics.*

*Software Used:*

*Python (pandas, numpy, matplotlib, seaborn, scikit-learn, xgboost)*

*Jupyter Notebook*

***Key Performance Indicators (KPIs)***

*These are the core metrics you analyzed to evaluate player and team performance:*

1. ***Goal Contributions (G+A)***
   * *Total goals + assists.*
   * *Primary measure of attacking output.*
2. ***Per 90 Metrics (Efficiency)***
   * *Goals per 90, Assists per 90, Goal+Assist per 90.*
   * *Allows fair comparison between players regardless of playing time.*
3. ***Expected Metrics (Quality of Chances)***
   * ***xG (Expected Goals)****: Measures quality of shots taken.*
   * ***xAG (Expected Assists)****: Measures quality of chances created.*
   * ***xG−Goals (Finishing Ability)****: Difference between goals scored and expected goals.*
4. ***Progression Metrics (Involvement in Build-up)***
   * ***Progressive Carries (PrgC)****: Times a player moves the ball significantly forward.*
   * ***Progressive Passes (PrgP)****: Forward passes advancing play.*
   * ***Carry/Pass Ratio****: Indicates playing style (ball carrier vs distributor).*
5. ***Discipline & Availability***
   * *Minutes played, matches started.*
   * *Yellow and red cards (affect availability).*

***Important Parameters for Modeling***

*When building predictive models for goal contributions, the following features mattered most:*

* ***Minutes played****: Strong volume factor — more time, more output.*
* ***xG and xAG****: Strongest predictors of goal contributions.*
* ***Progressive Carries & Passes****: Secondary factors influencing creativity.*
* ***Per-90 efficiency stats****: Help balance raw totals with efficiency.*
* ***Penalty Goals & Attempts****: Important for some players’ contributions (e.g., penalty-takers).*

***Final KPI Framework***

***Offensive Output*** *→ Goals, Assists, Goal+Assist, per-90 versions.*

* ***Chance Creation/Conversion*** *→ xG, xAG, xG Difference (Finishing).*
* ***Playmaking Style*** *→ Progressive Carries, Progressive Passes, Carry/Pass Ratio.*
* ***Volume & Reliability*** *→ Minutes, Matches Started, Cards.*

***Source:***

*The dataset used in this project was sourced from* ***FBref****, a football statistics platform powered by StatsBomb. It contains detailed player-level data for the Premier League 2024/25 season, including goals, assists, expected goals (xG), expected assists (xAG), minutes played, progression metrics, and disciplinary records. The raw data was made available via a community-uploaded dataset on* ***Kaggle****, which served as the basis for all cleaning, analysis, and modeling performed in this notebook.*

***Conclusion:***

*This project demonstrates that linear models best explain player contributions in the Premier League 2024/25 dataset. Ridge Regression proved most stable, highlighting that expected goals, expected assists, and minutes are the strongest predictors. The analysis shows that clubs and analysts should focus on per-90 efficiency alongside raw totals when evaluating talent.*